



Semantic Search and Discovery

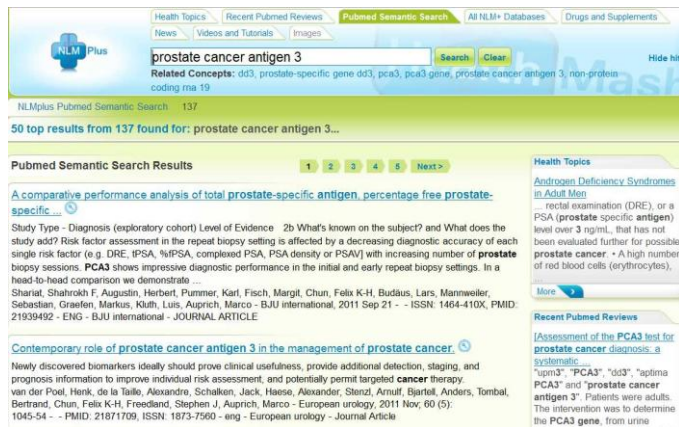
CENDI, January 12, 2012 @ LC

Tamas Doszkocs, Ph.D.

tamas@weblib.com

What is semantic search?

Keywords vs. Concepts



Health Topics Recent Pubmed Reviews Pubmed Semantic Search All NLM+ Databases Drugs and Supplements

News Videos and Tutorials Images

prostate cancer antigen 3

Related Concepts: d33, prostate-specific gene d33, pca3, pca3 gene, prostate cancer antigen 3, non-protein-coding rna 19

NLMplus Pubmed Semantic Search 137

50 top results from 137 found for: prostate cancer antigen 3...

Pubmed Semantic Search Results

1 2 3 4 5 Next>

A comparative performance analysis of total prostate-specific antigen, percentage free prostate-specific...

Study Type - Diagnosis (exploratory cohort) Level of Evidence 2b What's known on the subject? and What does the study add? Risk factor assessment in the repeat biopsy setting is affected by a decreasing diagnostic accuracy of each single risk factor (e.g. DRE, tPSA, %fPSA, complexed PSA, PSA density or PSAV) with increasing number of prostate biopsy sessions. PCA3 shows impressive diagnostic performance in the initial and early repeat biopsy settings. In a head-to-head comparison we demonstrate

Shariat, Shahrokh F, Augustin, Herbert, Pummer, Karl, Fisch, Margit, Chun, Felix K-H, Budius, Lars, Mannweiler, Sebastian, Graefen, Markus, Kluth, Luis, Aufrich, Marco - BJU international, 2011 Sep 21 - - ISSN: 1464-410X, PMID: 21939492 - ENG - BJU international - JOURNAL ARTICLE

Contemporary role of prostate cancer antigen 3 in the management of prostate cancer.

Newly discovered biomarkers ideally should prove clinical usefulness, provide additional detection, staging, and prognosis information to improve individual risk assessment, and potentially permit targeted cancer therapy

van der Poel, Henk, de la Taille, Alexandre, Schalken, Jack, Haese, Alexander, Slawik, Arnulf, Bjartell, Anders, Tombal, Bertrand, Chun, Felix K-H, Freedland, Stephen J, Aufrich, Marco - European urology, 2011 Nov, 60 (5): 1045-54 - - PMID: 21871709, ISSN: 1873-7560 - eng - European urology - Journal Article

Health Topics

Androgen Deficiency Syndromes in Adult Men

rectal examination (DRE), or a PSA (prostate specific antigen) level over 3 ng/ml, that has not been evaluated further for possible prostate cancer. • A high number of red blood cells (erythrocytes).

More

Recent Pubmed Reviews

Assessment of the PCA3 test for prostate cancer diagnosis: a systematic

"pca3", "PCA3", "d33", "aptima PCA3" and "prostate cancer antigen 3". Patients were adults. The intervention was to determine the PCA3 gene, from urine

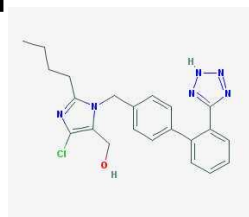
“**semantic search** is a search or a question or an action that produces **meaningful results**, even when the retrieved items contain **none of the query terms**, or the search involves **no query text** at all ”



What is NLMplus?

| NLM Plus | | | | | |
|--|---------|---|--------|--|-------|
| Health Topics Recent PubMed Reviews PubMed Semantic Search All NLM Databases Drugs and Supplements Feeds | | | | | |
| Voice and Tutorials Images | | | | | |
| <input type="text" value="hypertension"/> <input type="button" value="Search"/> <input type="button" value="Clear"/> <input type="button" value="Hide Info"/> | | | | | |
| All NLM+ Databases | | | | | |
| AID Info: HIV/AIDS Information | 10 | BioProject: aggregated biological research project data | 2 | Bio Sample: biological material descriptions | 1,177 |
| Books: online books | 6,566 | Cancer Chromosomes, cytogenetic databases | 0 | CCRB: Chemical Carcinogenesis Information | 1 |
| ChemIDplus: Chemical Identification Dictionary | 0 | ClinicalTrials.gov | 4,177 | CCDB: Conserved protein domain database | 22 |
| DART: Developmental and Reproductive Toxicology Database | 8264 | dbGFP: genotype and phenotype | 68,931 | CCDB: Comparative Toxicogenomics Database | 3211 |
| Drug Information Portal | 0 | dbGFP: genotype and phenotype | 68,931 | CCDB: Comparative Toxicogenomics Database | 3211 |
| GENETOX: Genetic Toxicology Data | 0 | dbGFP: genotype and phenotype | 68,931 | CCDB: Comparative Toxicogenomics Database | 3211 |
| GEO Profiles: expression and molecular abundance profiles | 0 | dbGFP: genotype and phenotype | 68,931 | CCDB: Comparative Toxicogenomics Database | 3211 |
| Household Products: Health & Safety Information on Household Products | Show me | dbGFP: genotype and phenotype | 68,931 | CCDB: Comparative Toxicogenomics Database | 3211 |
| ITER: International Toxicity Estimates for Risk | 1 | dbGFP: genotype and phenotype | 68,931 | CCDB: Comparative Toxicogenomics Database | 3211 |
| NLM Catalog: catalog of books, journals, and audio/visuals in the NLM collections | 4,734 | dbGFP: genotype and phenotype | 68,931 | CCDB: Comparative Toxicogenomics Database | 3211 |
| PopSet: population study data sets | 2 | dbGFP: genotype and phenotype | 68,931 | CCDB: Comparative Toxicogenomics Database | 3211 |
| PubChem BioAssay: bioactivity screens of chemical substances | 2,137 | dbGFP: genotype and phenotype | 68,931 | CCDB: Comparative Toxicogenomics Database | 3211 |
| PubMed: biomedical literature citations and abstracts | 336,037 | dbGFP: genotype and phenotype | 68,931 | CCDB: Comparative Toxicogenomics Database | 3211 |
| Structure: three-dimensional macromolecular structures | 324 | dbGFP: genotype and phenotype | 68,931 | CCDB: Comparative Toxicogenomics Database | 3211 |
| TRI: Toxics Release Inventory | 0 | dbGFP: genotype and phenotype | 68,931 | CCDB: Comparative Toxicogenomics Database | 3211 |
| | | dbGFP: genotype and phenotype | 68,931 | CCDB: Comparative Toxicogenomics Database | 3211 |

NLMplus is a **semantic search** and **knowledge discovery** application designed to tap into the **rich content offerings** of the **National Library of Medicine** in all areas of **biomedicine** and **health**



adhd aids alt amlodipine anemia anxiety arthritis aspirin asthma
atenolol atrial fibrillation autism bipolar blood pressure breast cancer bun calcium
cancer cbc celexa cellulitis cholesterol citalopram clonazepam clonidine colonoscopy
copd coumadin creatinine cystic fibrosis dementia depression diabetes digoxin
diverticulitis diverticulosis eczema endometriosis fibromyalgia furosemide
gabapentin gerd glaucoma gout heart hematocrit hemoglobin hepatitis c
hernia hiatal hernia hiv hydrochlorothiazide hyperlipidemia hypertension hypothyroidism
interactive clinical trial tutorial iron leukemia lipitor lisinopril liver
lorazepam lung cancer lupus magnesium melatonin metformin metoprolol mrsa
multiple sclerosis niacin obesity omeprazole osteoarthritis osteoporosis
pancreatitis pneumonia potassium prednisone pregnancy pyelonephritis
rheumatoid arthritis risperidone sarcoidosis scabies schizophrenia scoliosis seroquel
shingles simvastatin stroke synthroid thyroid tramadol vitamin c vitamin d
vitamin e zinc



NLMplus Background

- NLMplus is an [award winning](#) application of **semantic search** and **knowledge discovery** technology which was developed by [WebbLib LLC](#) during the first year of a [DOE/OSTI](#) SBIR project
- NLMplus addresses the difficult problem of **accessing heterogeneous legacy databases** with a meaningful solution in order to produce **improved relevance and precision in search results**
- The **goal** of the SBIR project was to research and implement **innovative semantic tools** for the **improved indexing, searching and dissemination of scientific and technical information.**
- As part of the SBIR research, [WebLib](#) had implemented a [Green Energy Semantic Search](#) pilot, which was subsequently deployed into [production use](#) by OSTI
- Where possible, WebLib's semantic solution **enriches** and utilizes **existing thesauri and ontologies** in a [Web Knowledge Base](#) which powers the semantic search.
 - Typical thesaurus problem: “**green energy**” is **NOT** in the ETDE/INIS Joint [Energy Thesaurus](#)

Semantic Annotation of PubMed Reviews

[anthracycline cardiotoxicity after breast cancer treatment](#)

 PubMed

US National Library of Medicine
National Institutes of Health

[Limits](#) [Advanced](#)

[Display Settings:](#) ☒ Abstract

[Send to:](#) ☒

[Oncology \(Williston Park\)](#). 2009 Mar;23(3):227-34.

Anthracycline cardiotoxicity after breast cancer treatment.

[Hershman DL](#), [Shao T](#).

Columbia University, New York, N Y 10032, USA.

Abstract

Anthracyclines are among the most active agents for the treatment of breast cancer; their use in combination regimens improves both disease-free and overall survival in patients with breast cancer. Unfortunately, the clinical utility of anthracycline use is limited by a cumulative dose-dependent cardiac toxicity resulting in congestive heart failure. As methods for detecting and treating breast cancer improve, there has been a steady decline in breast cancer mortality over the past 15 years. With an increasing number of long-term breast cancer survivors, the number of patients experiencing anthracycline-induced cardiotoxicity may also continue to grow. Moreover, new agents used in the treatment of breast cancer can potentiate cardiac toxicity. Recently, studies of non-anthracycline-containing regimens have been found to be effective in preventing recurrence of breast cancer (as compared with anthracycline-containing regimens) in patients with early-stage breast cancer, with a reduced incidence of adverse cardiac outcomes. In this article, we summarize the incidence, presentation, and mechanism of anthracycline-associated cardiotoxicity. We also discuss risk factors for the development of anthracycline-induced cardiotoxicity and new therapies, such as trastuzumab, that may potentiate cardiac toxicity. Finally, we review monitoring and preventive practices that may reduce the long-term risk of anthracycline-related cardiotoxicity.

Semantic Annotation of a PubMed Review Title

via the WebLib Part-of-Speech Tagger and Noun Phrase Chunker

phrases> **key phrases**> **key concepts** > **synonyms** > **normalized semantic indexes**

19418823;TI:anthracycline cardiotoxicity after breast cancer treatment

TI_NORMALP: **anthracycline breast canc cardiotoxicit treat**

TP:**anthracycline**;

- CUI:C0003234; Anthracycline Antibiotics;
- SY: anthracyclines;

– TP:**cardiotoxicity**;

– CUI:C0876994; **cardiac toxicity**;

– TP:**breast cancer**;

- CUI:C0678222; Breast Carcinoma;
- SY: breast tumor;

– TP:**breast cancer treatment**;

- CUI:C1511300; Breast Cancer Therapeutic Procedure;

Scaling the Semantic Searching of PubMed Reviews

- More than **1.6 million PubMed Reviews** have been **semantically indexed** on a small WebLib server
 - Query: prostate cancer antigen 3
 - Query: transient cortical blindness after coronary angiography

[Health Topics](#)[Pubmed Reviews Semantic Search](#)[Enhanced Pubmed Search](#)[All N](#)[Drugs and Supplements](#)[News](#)[Videos and Tutorials](#)[Images](#)[Search](#)[Clear](#)

Related Concepts: dd3, prostate-specific gene dd3, pca3, pca3 gene, prostate cancer a
coding ma 19

[Latest Reviews from Pubmed](#)

0

[NLMplus Pubmed Reviews Semantic Search](#)

40 (from 40)

40 top results from 40 found for: prostate cancer antigen 3...

Pubmed Reviews Semantic Search Results

[1](#)[2](#)[3](#)[4](#)[Next >](#)

[\[Assessment of the PCA3 test for prostate cancer diagnosis: a systematic ...\]](#)

"upm3", "PCA3", "dd3", "aptima PCA3" and "prostate cancer antigen 3". Patients were adults. The intervention was to determine the **PCA3 gene**, from urine samples for diagnosis of **prostate cancer**. The quality ... OBJECTIVE: To assess the efficacy of diagnostic techniques based on **PCA3 gene** for early detection
[Ruiz-Aragon J, Marquez-Pelaez S] - [Actas Urol Esp, 2010; 34(4):346-355]


[PCA3: from basic molecular science to the clinical lab.](#)

are emerging and **prostate cancer gene 3 (PCA3)** is one such marker. **PCA3** is a noncoding RNA that is highly over-expressed in **prostate cancer** tissue compared to benign tissue. A non-invasive test for **PCA3** was developed ... demonstrated the utility of **PCA3** for the diagnosis of **prostate cancer** and some studies suggest that **PCA3** may
[Day JR, Jost M, Reynolds MA, Groskopf J, Rittenhouse H] - [Cancer Lett., 2011; 301(1):1-6]

[Progensia PCA3 test for prostate cancer detection.](#)

are required in order to avoid unnecessary biopsies. The **PCA3 gene** product is specifically overexpressed ... in order to

Search Results from NLM's PubMed Search Engine



US National Library of Medicine
National Institutes of Health

PubMed

prostate cancer antigen 3

Search

RSS Save search Limits Advanced

Help

Display Settings: ☒ Summary, 20 per page, Sorted by Recently Added

Send to: ☒ Filter your results:

Results: 1 to 20 of 909

<< First < Prev Page 1 of 46 Next > Last >>

☐ [Antitumor effects of bisphosphonates: from the laboratory to the clinic.](#)

1. Berenson JR.
Curr Opin Support Palliat Care. 2011 Sep;5(3):233-40. Review.
PMID: 21825999 [PubMed - indexed for MEDLINE]
[Related citations](#)

☐ [The complex interplay between cholesterol and prostate malignancy.](#)

2. Solomon KR, Freeman MR.
Urol Clin North Am. 2011 Aug;38(3):243-59. Epub 2011 Jun 22. Review.
PMID: 21798387 [PubMed - indexed for MEDLINE]
[Related citations](#)

☐ [Carbohydrate molecules: an expanding horizon in drug delivery and biomedicine.](#)

3 Shukla RK, Tiwari A.
Crit Rev Ther Drug Carrier Syst. 2011;28(3):255-92. Review.

All (11009)

Free Full Text (2456)

Review (909)

[Manage Filters](#)

Find related data

Database: Select

Find items

Search details

("prostatic neoplasms"[MeSH Terms] OR ("prostatic"[All Fields] AND "neoplasms"[All Fields]) OR "prostatic neoplasms"[All Fields] OR

4 top results from 4 found for: transient cortical blindness after coronary angiography...

Pubmed Reviews Semantic Search Results

[Transient cortical blindness after coronary angiography: a ...](#)

Transient cortical blindness after coronary angiography is a rarely encountered, rapidly ... study. We report a case of **transient cortical blindness** following diagnostic **coronary angiography** ... include those undergoing **coronary** artery bypass graft study and those with renal failure. Although
[Akhtar N, Khatri IA, Naseer A, Ikram J, Ahmed W] - [J Pak Med Assoc, 2011; 61(3):295-297]

[Cortical blindness: a rare but dramatic complication following coronary angiography.](#)

[Gellen B, Remp T, Mayer T, Milz P, Franz WM] - [Cardiology, ; 99(1):57-59]

[An uncommon complication after a common procedure.](#)

A 50-year-old female developed **cortical blindness** 5 hours **after coronary angiography**. A non-contrast-enhanced computed tomographic scan of the brain revealed bilateral occipital and middle cerebral/posterior cerebral artery watershed territory infarcts with a possibility of subarachnoid hemorrhage (SAH). Magnetic
[Math RS, Singh S, Bahl V] - [J Invasive Cardiol, 2008; 20(10):E301-E303]

[Transient cortical blindness following bypass graft angiography. A case report.](#)

Transient cortical blindness, an uncommonly recognized complication of cerebral **angiography** ... -old patient who had **transient cortical blindness** following bypass graft **angiography**. In this case ... the volume injected. When **cortical blindness** occurs, it is a serious clinical problem whether **transient**
[Kamata J, Fukami K, Yoshida H, Mizunuma Y, Moriai N, Takino T, Hosokawa S, Hashimoto K, Nakai K, Kawazoe K, et al] - [Angiology, 1995; 46(10):937-946]

PubMed Query: **transient cortical blindness following coronary angiography**

 PubMed 

US National Library of Medicine
National Institutes of Health

 [RSS](#) [Save search](#) [Limits](#) [Advanced](#)

Display Settings: ☒ Summary, Sorted by Recently Added

Send to:

Results: 2

- ☐ [Transient cortical blindness after coronary angiography: a case report and literature review.](#)
 1. Akhtar N, Khatri IA, Naseer A, Ikram J, Ahmed W.
J Pak Med Assoc. 2011 Mar;61(3):295-7. Review.
PMID: 21465952 [PubMed - indexed for MEDLINE]
[Related citations](#)
- ☐ [Transient cortical blindness following bypass graft angiography. A case report.](#)
 2. Kamata J, Fukami K, Yoshida H, Mizunuma Y, Moriai N, Takino T, Hosokawa S, Hashimoto K, Nakai K, Kawazoe K, al.
Angiology. 1995 Oct;46(10):937-46. Review.
PMID: 7486215 [PubMed - indexed for MEDLINE]
[Related citations](#)

The WebLib Semantic Search Engine

- Apache Solr open source enterprise search engine with WebLib's Semantic Layer
 - **Please note:** WebLib's semantic indexing and search technology can be implemented for any enterprise search engine, e.g. the MarkLogic system, which is being increasingly utilized by federal agencies
- WebLib's Biomedical Knowledge Base and Query Translator
- WebLib's Web Knowledge Base Prototype

More about the NLMplus

NLMplus Innovations

- Universal Search
 - Flexible access to 60 NLM databases
- Semantic Search Engine
 - Semantic indexing and searching of PubMed Reviews
- Biomedical Knowledge Base
 - powers “Explore and Discover”
 - powers the Semantic Search Engine
- Intuitive User Interface
- Wide Spectrum of NLM users
 - Consumers
 - Physicians
 - Biomedical researchers
 - Decision makers

NLM APIs, web services and software tools utilized by the NLMplus app

- [Entrez Programming Utilities](#)
- [UMLS Terminology Services](#)
- [Medlineplus Web Service](#)
- [MetaMap API](#)
- [SKR Web API](#)

Potential WebLib Semantic Search and Knowledge Base Applications

Common Problem: *accessing heterogeneous databases*

Problems in Keyword Searching: *relevance and precision*

Meaningful Solution: *semantic technologies*

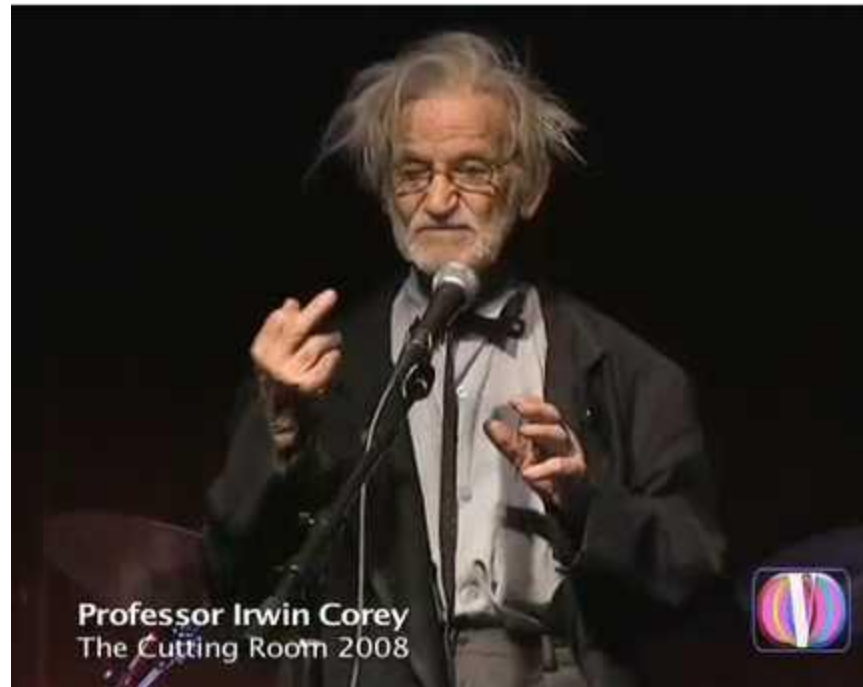
Biomedical Knowledge Base Semantic Search

- National Library of Medicine
- National Institutes of Health
- Department of Health and Human Services
- Medical Schools
- Health Science Libraries
- Health Care Industry

Web Knowledge Base Semantic Search

- U.S. Government Departments and Agencies
- Libraries and Research Organizations
- Media and Content Companies
- Educational Institutions
- Businesses
- The Web

Professor Irwin Corey Explains Semantics





Semantic Search and Discovery

CENDI, January 12, 2012 @ LC

Tamas Doszkocs, Ph.D.

tamas@weblib.com